Can cognitive science help us make information risk more tangible online?

Sadie Creese
International Digital Laboratory
The University of Warwick

Koen Lamberts
Centre for Cognitive and Neural Systems
The University of Warwick

Individuals' ability to assess the risks associated with operating in cyberspace has potentially important implications for themselves and for the social and corporate networks to which they belong. Failing to detect fraudulent email or 'phishing' attempts, accidentally revealing confidential or sensitive information on social networking sites, or inadvertently downloading malicious software, are all examples of online risks that are faced by every user of the Web. The ever-growing pervasive and promiscuous connectivity of digital information devices offers many more chances for malware spread and unauthorised access to personal and corporate facilities, potentially compromising confidentiality, integrity and availability of personal data, devices, corporate networks and critical infrastructures. The consequences of failing to detect and correctly handle online threats can be extremely serious (e.g., Poulsen, 2003). Some online threats are intrinsically difficult to manage, especially those that involve direct communication between users (on social networking sites or online chatrooms). Without direct monitoring of such interactions, it is impossible to detect manipulative or deceptive behavior, and users have to rely on everyday common sense to maintain their safety and security. This can be difficult, especially in contexts where nonverbal cues are not available, as in chatrooms without video links (e.g., Burgoon, Blair, & Strom, 2008). However, there also risky situations that do not involve direct interaction with another user (such as the risk of downloading malicious software), and because these situations tend to have a much more tractable structure than direct user-to-user contact, they offer better prospects for controlled risk management. In this article, we will focus primarily on risk contexts of the latter kind.

Although individuals are equipped with cognitive tools that allow them to assess risk in a wide range of situations and contexts (e.g., Slovic, 1987), much evidence suggests that people do not find risk in cyberspace a tangible concept (e.g., Jackson et al., 2005). Whereas everybody knows that locking a door can prevent unauthorized entry, people tend not to understand the equivalent precautions they can take in order to protect their data and communications devices in cyberspace. The general view of the security professionals' community is that this is a growing problem, as the incidence of cyber attacks exploiting human vulnerabilities and identity theft involving digital credentials is increasing (see McAfee, 2008).

Most online interfaces contain tools that are designed to help users protect their information, such as images of virtual padlocks, pop-up windows warning of potential for vulnerability, and threat indicators that give a numerical representation of the likelihood and consequences of particular adverse events. However, the growing incidence of security failures suggests that these tools are not entirely effective. The reasons for this are poorly understood, and clearly merit further research. However, it is not difficult to see the potential limitations of current visual aids and threat indicators. Risk perception is ultimately perception of the likelihood and the consequences of uncertain events. Therefore, information risk not only depends on the existence and nature of a threat and the presence of a vulnerability that such a threat seeks to exploit, but also on the impact of a possible adverse event. Usually, this impact can only be judged by the owner of the information or the system at risk, because the cost of loss of integrity, availability or confidentiality of information tends to be subjective and context dependent. Existing research on usable security does consider human users as central to the design of security solutions (e.g., Balfanz et al., 2004; Dourish & Redmiles, 2002; Rode et al., 2006), but research has generally focused on pragmatic solutions for specific problems, without offering a general theoretical framework that is widely applicable. There is currently no agreed set of design principles for

a system that supports accurate communication and assessment of online risk, and that supports users in choosing the appropriate action in a given context. Whereas education and self-help material is widely available and usually of good quality, and commercial initiatives designed to encourage uptake of security practice are commendable, the fact that users continue to take actions resulting in undesired harm to themselves demonstrates the need for more effective solutions. Web users need to be provided with a better understanding of the risks they are taking in a range of contexts, by using effective tools with high usability.

Whereas risk has been studied in several disciplines (including economics, political science, biology and anthropology), empirical research in cognitive psychology is perhaps most relevant for understanding the principles that determine people's perception and assessment of risk. The extensive literature on the cognitive processes that people use to assess risk (e.g., Gigerenzer & Goldstein, 1996; Kahneman, Slovic, & Tversky, 1982; Slovic, 1987) has shown that risk perception and assessment are complex processes, that can involve the application of various heuristics (such as representativeness and availability; see Kahneman et al., 1982) that can lead to sub-optimal evaluation. Moreover, risk perception can be based on distorted or incomplete representations of the environment, which can introduce further distortions. Research has also highlighted the complex relations between trust, trustworthiness and risk (e.g., Colquitt et al., 2007). Methodological complications follow from the fact that risk assessment is often an implicit (unconscious) process, which implies that people have poor introspective access to the mental operations they use to evaluate risk and make decisions about actions.

Although existing research in cognitive psychology can undoubtedly provide important principles for the design of effective risk communication strategies, empirical research will have to indicate whether these principles can be applied to Web interfaces. Many aspects of activities in cyberspace are not well documented, and there is no generally accepted account of how people represent and process information in virtual environments. What is required is a comprehensive theoretical framework for communication and interaction on the Web that articulates entire processing sequences, which include (i) initial perception of information, (ii) construction of a representation of the virtual environment, including assessment of any risks present, (iii) decision making and subsequent action. Such a framework will support the design of interfaces that better communicate the level and nature of risk that users are exposed to. In this article, we will not attempt to provide such a framework – too many aspects of relevant behavior and cognition are still poorly understood. Instead, we will focus on an issue that has been explored extensively in other contexts (such as health psychology), and explore how it might be applied to online settings: Which communication tools can be used to make risk as tangible as possible?

Using visual tools to represent risk

It is well established that graphical or pictorial representations can be used to improve understanding of risk (e.g., Weinstein & Sandman, 1993). Graphical representations can be used to emphasize particular aspects of the data, to reveal regularities that would otherwise be difficult to spot, to highlight particular relationships, and so forth (e.g., Lipkus & Hollands, 1999). In the context of risk assessment, graphical tools are particularly relevant, because they can be used to convey information about a number of different risk conditions or parameters in a direct, intuitively accessible manner. There is an extensive body of research in cognitive psychology about the effectiveness of various kinds of pictorial representation for different purposes, although most of the work remains largely atheoretical. In the context of risk communication, two considerations seem particularly relevant. First, graphical representations of risk magnitude should be consistent with people's mental representation of magnitudes. People find it difficult to represent absolute magnitudes in an accurate way, and instead

tend to rely on relative magnitudes wherever possible (e.g., Stewart, Brown, & Chater, 2005). Therefore, a graphical representation that emphasizes relative risk magnitudes is often most effective. This principle has been used successfully in so-called risk ladder representations, in which a target risk is placed on the same scale as other, perhaps more familiar, risks. The information that is derived from these anchoring points greatly enhances accurate risk assessment. Second, the visual information that is conveyed should be tailored to the task at hand (Sparrow, 1989). The representation that is used should facilitate the operations that are required to make the optimal judgment in the given context (Carswell, 1992). For instance, if the user needs to decide from which site to download a particular software package, an appropriate representation would allow an accurate comparison of infection risks from the different download sites. This can be achieved best by a graphical representation that emphasizes, and possibly exaggerates, the relative risk position of the choice alternatives, even if this means that absolute risk assessment becomes more difficult; the main concern in this context is to guide the user towards the lowest-risk option. A detailed discussion of visual tools for risk communication in various contexts can be found in Lipkus and Hollands (1999).

Numerical representation of risk

Naïve observers are not very good at reasoning about uncertain events, and often rely on imperfect heuristics to achieve an understanding of the likelihood of particular outcomes. Wherever this occurs, it is important that information is presented in a way that either (i) facilitates the use of appropriate heuristics, or (ii) discourages reliance on inappropriate heuristics. This principle becomes particularly important in circumstances where naïve reasoning can lead to serious errors. A good example can be found in reasoning about conditional probabilities (e.g., Gigerenzer & Edwards, 2003). To illustrate the issue, we carried out a simple study, in which we first presented a sample of 30 university students (from different backgrounds, but excluding students of mathematics, statistics and economics) with the following problem:

> *You want to make a purchase using your credit card on a website that you have not visited before. However, as you are about to enter your credit card details and other personal information, a warning message appears on the screen, to say that your antivirus software has detected code on the website that can be used by criminal hackers to conceal a virus that logs your personal details, and transfers them to the hackers' personal computers (who can then use them to steal your identity, for illegal purposes). 0.1% of all commercial websites that accept credit cards are infected with the virus. If a website is infected, the probability that your antivirus software will show a warning message is 99.9%, which means that there is a 0.1% probability that an infection will be missed by the antivirus software. If a website is not infected, the probability that a warning message will be shown (i.e., a false alarm) is 0.1%, and the probability that no warning message will be shown is 99.9%. What is the probability that the website on which you are about to enter sensitive information is infected, considering that you got a warning message from your antivirus software?*

The correct answer to this problem is given by Bayes' theorem, as follows (*i* = infected; *w* = warning given):

$$(\mid) = \frac{(\mid) ( )}{(\mid) ( ) + (\mid) ( )} = \frac{0.999 \times 0.001}{0.999 \times 0.001 + 0.001 \times 0.999} = 0.5$$

Of the 30 students, only three produced the correct answer. 16 participants incorrectly answered 99% or 99.9%, one participant answered 0.1%, and the others either failed to produce a response or gave yet a different value. Presented in this form, the problem is clearly beyond the grasp of most naïve subjects.

It is possible, however, to present the same information in a format that is intuitively accessible, and that supports a correct evaluation of the security risks presented. If the probabilities are translated into natural frequencies, assessment of risk can be much improved (e.g., Gigerenzer & Edwards, 2003; Gigerenzer & Hoffrage, 1995). Natural frequencies are simple counts of events in some defined context, and because they are not computed relative to different reference classes (as conditional probabilities are) they are much easier to interpret. As an example, consider a natural-frequency equivalent of the virus alert problem:

> *Out of every 10,000 commercial websites that accept credit cards, 10 are infected with a virus. These 10 infected websites will all produce a warning message from your antivirus software when you visit them. Of the 9990 websites that are not infected by the virus, 10 will produce a warning message from your antivirus software (false alarms), and 9,980 will not produce a warning message. Now think of 20 cases in which a virus warning message has been produced. In how many of those cases would you expect a virus be present?*

This description was presented to another sample of students. This time, 13 students out of 30 were able to give the correct response, without any reference to Bayes' theorem, usually by producing the correct warning/infection contingency table in some form (see Table 1). The removal of confusion about the reference classes to which the conditional probabilities refer clearly made a significant difference to people's ability to assess the presented risks.

	Infected	Not infected
Detected	10	10
Not detected	0	9,980

Table 1. Contingency table for the relation between virus infection and detection

Conclusions

Despite their limitations, the examples that we provided illustrate the importance of the format in which risk information is presented. Representations that are logically or numerically equivalent can still induce very different reasoning processes. The main lesson to be learnt from the cognitive psychology of risk communication is that understanding the link between aspects of representations and the mental processes they afford is critical for the design of effective risk communication. Although this principle is hardly novel (and has always been a cornerstone of cognitive ergonomics), it has not been sufficiently recognized in the context of online risk communication. The examples that we gave provide an illustration of how a cognitive framework can guide empirical research, which can eventually lead to the development of more effective tools for online risk assessment and management. We anticipate that a range of existing information and network security techniques, such as threat warnings and security posture indicators on web browsers, privacy controls and virus software, could be greatly enhanced simply by helping users deploy them more effectively. Our future research will be focused on the development of a comprehensive theoretical framework for communication and interaction on the Web that articulates entire processing sequences enabling us to understand risk taking, and the exploitation

of that framework to design new interfaces for information security techniques in order to enhance usability and ultimately security.

References

Balfanz, D., Durfee, G., Smetters, D.K., & Grinter, R.E. (2004). In search of usable security: Five lessons from the field. *Security & Privacy, IEEE*, 2, 19-24.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34,* 572-599.

Carswell, C. M. (1992). Choosing Specifiers - an Evaluation of the Basic Tasks Model of Graphical Perception. *Human Factors, 34*(5), 535-554.

Colquitt, J., Scott, B., & LePine, J. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology, 92*, 909-927.

Dourish, P. & Redmiles, D. (2002). An approach to usable security based on event monitoring and visualization. *Proceedings of the 2002 workshop on New security paradigms*, Virginia Beach, Virginia.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal, 327,* 741-744.

Gigerenzer, G. & Goldstein, D. (1996) Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review, 103*, 650-669.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Jackson, J., Allum, N. and Gaskell, G. (2005). Perceptions of Risk in Cyber Space, In: R. Mansell, & R. Collins, (Eds.), *Trust and Crime in Information Societies*. Edward Elgar, London.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute Monographs, 25,* 149-163.

McAfee (2008) The McAfee Global Threat Report 2008. Online publication: http://www.mcafee.com/uk/local_content/reports/gtr_2008.html

Poulsen, K. (2003). Slammer worm crashed Ohio nuke plant network. Online publication: http://www.securityfocus.com/news/6767

Rode, J., et al. (2006). Seeing further: extending visualization as a basis for usable security. *ACM International Conference Proceeding Series, 149*, 145-155.

Slovic, P. (1987). Perception of risk. *Science, 236*, 280-285.

Sparrow, J. A. (1989). Graphical displays in information systems: Some data properties influencing the effectiveness of alternative forms. *Behaviour & Information Technology, 8*(1), 43-56.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112*(4), 881-911.

Weinstein, N. D., & Sandman, P. M. (1993). Some criteria for evaluating risk messages. *Risk Analysis, 13*, 103-114.